



智能科学技术著作丛书

机器学习理论、方法及应用

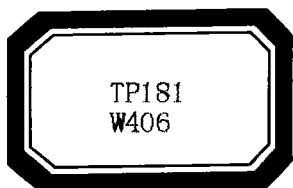
王雪松 程玉虎 著

81
6



科学出版社

www.sciencep.com



智能科学技术著作丛书

机器学习理论、方法及应用

王雪松 程玉虎 著

国家自然科学基金资助项目(60804022)

教育部“新世纪优秀人才支持计划”项目(NCET-08-0836)

江苏省自然科学基金资助项目(BK2008126)

教育部高等学校博士学科点专项科研基金资助项目(20070290537,
200802901506)

TP181
W406

科学出版社

北京

内 容 简 介

机器学习的研究不仅是人工智能领域的核心问题,而且已成为近年来计算机科学与技术领域中最活跃的研究分支之一。

本书主要围绕基于神经网络的学习、强化学习和进化学习三个方面阐述机器学习理论、方法及其应用,共三部分 13 章。第一部分是神经网络学习及其在复杂非线性系统中的控制,包括基于时间差分的神经网络预测控制,基于径向基函数网络的机械手迭代学习控制,自适应 T-S 型模糊径向基函数网络等。第二部分是强化学习的大规模或连续空间表示问题,包括基于强化学习的自适应 PID 控制,基于动态回归网络的强化学习控制,基于自适应模糊径向基函数网络、支持向量机和高斯过程的连续空间强化学习,基于图上测地高斯基函数的策略迭代强化学习等。第三部分则是对分布估计优化算法进行研究,包括多目标优化问题的差分进化-分布估计算法,基于细菌觅食行为的分布估计算法在预测控制中的应用,一种多样性保持的分布估计算法及其在支持向量机参数选择问题中的应用等。为便于应用本书阐述的算法,书后附有部分机器学习算法 MATLAB 源程序。

本书可供理工科高等院校计算机科学、信息科学、人工智能和自动化技术及相关专业的教师及研究生阅读,也可供自然科学和工程技术领域中的研究人员参考。

图书在版编目(CIP)数据

机器学习理论、方法及应用/王雪松,程玉虎著. —北京:科学出版社,2009
(智能科学技术著作丛书)

ISBN 978-7-03-025439-9

I. 机… II. ①王…②程… III. 机器学习 IV. TP181

中国版本图书馆 CIP 数据核字(2009)第 153909 号

责任编辑:姚庆爽 / 责任校对:赵桂芬

责任印制:赵 博 / 封面设计:陈 敬

科学出版社出版

北京东黄城根北街 16 号

邮政编码:100717

<http://www.sciencep.com>

数字印刷厂印刷

科学出版社发行 各地新华书店经销

*

2009 年 8 月第 一 版 开本:B5(720×1000)

2009 年 8 月第一次印刷 印张:12 1/4

印数:1—3 000 字数:224 000

定价:40.00 元

(如有印装质量问题,我社负责调换(环伟))

《智能科学技术著作丛书》序

“智能”是“信息”的精彩结晶，“智能科学技术”是“信息科学技术”的辉煌篇章，“智能化”是“信息化”发展的新动向、新阶段。

“智能科学技术”(intelligence science & technology, IST)是关于“广义智能”的理论方法和应用技术的综合性科学技术领域,其研究对象包括:

- “自然智能”(natural intelligence, NI),包括:“人的智能”(human intelligence, HI)及其他“生物智能”(biological intelligence, BI)。
- “人工智能”(artificial intelligence, AI),包括:“机器智能”(machine intelligence, MI)与“智能机器”(intelligent machine, IM)。
- “集成智能”(integrated intelligence, II),即:“人的智能”与“机器智能”人机互补的集成智能。
- “协同智能”(cooperative intelligence, CI),指:“个体智能”相互协调共生的群体协同智能。
- “分布智能”(distributed intelligence, DI),如:广域信息网,分散大系统的分布式智能。

1956年,“人工智能”学科诞生,50年来,在起伏、曲折的科学征途上不断前进、发展,从狭义人工智能走向广义人工智能,从个体人工智能到群体人工智能,从集中式人工智能到分布式人工智能,在理论方法研究和应用技术开发方面都取得了重大进展。如果说,当年“人工智能”学科的诞生是生物科学技术与信息科学技术、系统科学技术的一次成功的结合,那么,可以认为,现在“智能科学技术”领域的兴起是在信息化、网络化时代又一次新的多学科交融。

1981年,“中国人工智能学会”(Chinese Association for Artificial Intelligence, CAAI)正式成立,25年来,从艰苦创业到成长壮大,从学习跟踪到自主研发,团结我国广大学者,在“人工智能”的研究开发及应用方面取得了显著的进展,促进了“智能科学技术”的发展。在华夏文化与东方哲学影响下,我国智能科学技术的研究、开发及应用,在学术思想与科学方法上,具有综合性、整体性、协调性的特色,在理论方法研究与应用技术开发方面,取得了具有创新性、开拓性的成果。“智能化”已成为当前新技术、新产品的发展方向和显著标志。

为了适时总结、交流、宣传我国学者在“智能科学技术”领域的研究开发及应用成果,中国人工智能学会与科学出版社合作编辑出版《智能科学技术著作丛书》。需要强调的是,这套丛书将优先出版那些有助于将科学技术转化为生产力以及对社会和国民经济建设有重大作用和应用前景的著作。

我们相信,有广大智能科学技术工作者的积极参与和大力支持,以及编委们的共同努力,《智能科学技术著作丛书》将为繁荣我国智能科学技术事业、增强自主创

新能力、建设创新型国家做出应有的贡献。

祝《智能科学技术著作丛书》出版,特赋贺诗一首:

**智能科技领域广
人机集成智能强
群体智能协同好
智能创新更辉煌**

涂序彦

中国人工智能学会荣誉理事长

2005年12月18日

序

机器学习是人工智能领域最早研究的内容之一,也是最体现智能特征的研究方向之一,不仅具有深刻的理论内涵,而且是现代社会中人们获取和处理知识的重要技术手段。随着机器学习研究的不断深入和计算机技术的长足进步,已经设计出不少具有优良性能的机器学习系统并投入实际应用。机器学习已经成为当今人工智能中最受关注的领域之一。

该书是作者近年来在国家自然科学基金、教育部“新世纪优秀人才支持计划”、江苏省自然科学基金以及教育部高等学校博士学科点专项科研基金项目的资助下,取得的一系列关于机器学习研究成果的结晶,不仅对已有研究成果做了全面总结,也是对当前机器学习研究成果的重要补充。书中全面、系统地介绍了机器学习的基本概念、发展历史、分类及部分机器学习的主要策略等,并重点围绕当前机器学习领域的热点问题展开讨论,包括:神经网络学习及其在复杂非线性系统控制中的应用、大规模或连续空间下的强化学习以及分布估计优化算法等。此外,为理论联系实际和便于读者理解算法思想,书中还介绍了机器学习方法的若干典型应用,如机械手轨迹跟踪控制、小车爬山最短时间控制、倒立摆平衡控制、小船过河控制、机器人迷宫行走以及复杂数值函数优化问题等。

该书体现的学术思想新颖且内容广泛,理论联系实际,写作结构清晰,逻辑性强,阐述严谨。相信该书的出版一定能进一步推动和促进机器学习领域的研究与发展,为广大有意致力于机器学习研究与应用的人们提供一本难得的参考书。

费树岷

东南大学自动化学院

前 言

学习是人类具有的一种重要智能行为,而机器学习是一门研究怎样用计算机来模拟或实现人类学习活动的学科。机器学习从很多学科吸收了成果和概念,包括人工智能、概率论与数理统计、哲学、信息论、生物学、认知科学和控制论等,是多门学科有机交叉的新颖研究方向。机器学习的研究不仅是人工智能领域的核心问题,而且已成为近年来计算机科学与技术领域中最活跃的研究分支之一。

随着人工智能研究的进展,人们逐渐发现研究人工智能的最好方法是向人类自身学习,因而引入了一些模拟进化的方法来解决复杂优化的问题,其中富有代表性的是遗传算法。遗传算法的生物基础是人类生理的进化及发展,这种方法被称为进化主义;另一方面,神经网络的理论是基于人脑的结构,其目的是揭示一个系统是如何向环境学习的,此方法被称为联结主义。这两种方法与传统机器学习方法大相径庭,因而近年来许多科学家致力于这两种方法的研究。另外,美国心理学家、动物心理学的开创者、心理学联结主义的建立者和教育心理学体系的创始人Thorndike关于动物学习心理学的研究表明,从挫折和失败中积累经验和知识不但是人类,也是高等哺乳动物学习的重要方式,心理学家称之为“试错法”学习。这种“试错法”的思想反映在机器学习中,就是强化学习。在20世纪80年代,基于试错法、动态规划和时间差分方法形成了强化学习理论,这是一种不同于传统机器学习理论的学习方法。因此,基于神经网络的学习、进化学习以及强化学习就构成了机器学习的三个主要策略。

作者长期从事机器学习的研究工作,在国家自然科学基金、教育部“新世纪优秀人才支持计划”、江苏省自然科学基金以及教育部高等学校博士学科点专项科研基金项目资助下,提出了一整套机器学习理论,解决了一系列机器学习方法的核心技术,并将其成功地应用于许多复杂的实际问题。这些工作大大丰富了机器学习理论,提高了机器学习方法解决实际问题的能力,也为机器学习方法在其他领域的进一步应用奠定了技术基础,具有重要的理论意义和实际应用价值。

本书是作者在国内外本领域权威期刊以及有影响的国际会议论文集上发表的十余篇学术论文的基础上进一步加工、深化而成的,是对已有研究成果的全面总结。本书主要围绕基于神经网络的学习、强化学习和进化学习三个方面阐述机器学习理论、方法及其应用,共三部分13章。第一部分是神经网络学习及其在复杂非线性系统中的控制,包括:基于时间差分的神经网络预测控制,基于径向基函数网络的机械手迭代学习控制,自适应T-S型模糊径向基函数网络。第二部分是强

化学习的大规模或连续空间表示问题,包括:基于强化学习的自适应 PID 控制,基于动态回归网络的强化学习控制,基于自适应模糊径向基函数网络、支持向量机和高斯过程的连续空间强化学习,基于图上测地高斯基函数的策略迭代强化学习。第三部分则是对分布估计优化算法进行研究,包括:多目标优化问题的差分进化-分布估计算法,基于细菌觅食行为的分布估计算法在预测控制中的应用,一种多样性保持的分布估计算法及其在支持向量机参数选择问题中的应用。作者愿将这些研究成果与国内外同行一起分享,以推动该领域的进一步研究与发展。

本书的撰写参考了国内外有关研究成果,他们的丰硕成果和贡献是本书学术思想的重要源泉,在此对涉及的专家和研究人员表示衷心的感谢。作者得到东南大学博士生导师费树岷教授多方面的指导,费树岷老师在百忙之中不但仔细地审阅了全部书稿,提出了许多非常中肯的建议和意见,而且欣然为本书作序,令作者深受鼓舞,在此向费树岷老师表示衷心的感谢!中国矿业大学信息与电气工程学院的马小平教授、李明教授、孙伟教授、巩敦卫教授等为本书的撰写提供了许多有益的指导,江苏省重点学科“控制理论与控制工程”博士点也为本书的出版提供了大力的帮助,在此一并表示感谢。

机器学习是一个快速发展、多学科交叉的新颖研究方向,其理论及应用均有大量的问题尚待进一步深入地研究。由于作者学识水平和可获得资料的限制,书中尚有不妥之处,敬请同行专家和读者批评指正。

作 者

2009年4月于中国矿业大学

目 录

《智能科学技术著作丛书》序

序

前言

第 1 章 机器学习概述	1
1.1 机器学习的概念	1
1.2 机器学习的发展历史	3
1.3 机器学习的分类	4
1.3.1 基于学习策略的分类	4
1.3.2 基于学习方法的分类	4
1.3.3 基于学习方式的分类	4
1.3.4 基于数据形式的分类	5
1.3.5 基于学习目标的分类	5
1.4 机器学习的主要策略	5
1.4.1 基于神经网络的学习	6
1.4.2 进化学习	9
1.4.3 强化学习	14
1.5 本书主要内容及安排	16
1.6 本章小结	17
参考文献	18
第 2 章 基于时间差分的神经网络预测控制	22
2.1 方法的提出	23
2.2 基于时间差分的 Elman 网络预测控制	24
2.2.1 Elman 网络预测模型	24
2.2.2 反馈校正模型	26
2.2.3 参考轨迹	27
2.2.4 滚动优化算法	27
2.3 仿真研究	28
2.3.1 预测仿真	28
2.3.2 跟踪仿真	29
2.4 本章小结	30

参考文献	30
第 3 章 基于径向基函数网络的机械手迭代学习控制	31
3.1 机械手迭代学习控制	32
3.2 基于 RBF 网络的迭代学习控制	32
3.2.1 选取查询点的 k 个最接近样例	33
3.2.2 利用 RBF 网络拟合 k 个数据点	33
3.2.3 预测查询点的控制输入	34
3.3 仿真研究	35
3.4 本章小结	36
参考文献	36
第 4 章 自适应 T-S 型模糊径向基函数网络	38
4.1 RBF 网络和模糊推理系统的功能等价性	39
4.2 自适应 T-S 型 FRBF 网络结构	40
4.3 自适应 T-S 型 FRBF 网络学习	41
4.3.1 网络学习动态	42
4.3.2 网络结构学习	42
4.3.3 网络参数学习	44
4.3.4 算法步骤	45
4.4 仿真研究	45
4.5 本章小结	47
参考文献	47
第 5 章 基于强化学习的自适应 PID 控制	48
5.1 Actor-Critic 学习	49
5.2 基于强化学习的自适应 PID 控制	50
5.2.1 基于强化学习的自适应 PID 控制结构	50
5.2.2 基于 RBF 网络的 Actor-Critic 学习	51
5.3 控制器设计步骤	53
5.4 仿真研究	53
5.5 本章小结	54
参考文献	55
第 6 章 基于动态回归网络的强化学习控制	56
6.1 Q 学习	56
6.2 基于 Elman 网络的强化学习控制	57
6.2.1 基于 Elman 网络的 Q 学习	57
6.2.2 Elman 网络学习算法	59

6.2.3 基于 Elman 网络的 Q 学习方法步骤	60
6.3 仿真研究	60
6.4 本章小结	62
参考文献	63
第 7 章 基于自适应 FRBF 网络的强化学习	64
7.1 基于自适应 FRBF 网络的 Actor-Critic 学习	64
7.1.1 基于自适应 FRBF 网络的 Actor-Critic 学习结构	64
7.1.2 自适应 FRBF 网络的学习	65
7.1.3 算法步骤	67
7.1.4 仿真研究	68
7.2 基于自适应 FRBF 网络的 Q 学习	72
7.2.1 基于自适应 FRBF 网络的 Q 学习结构	72
7.2.2 自适应 FRBF 网络的学习	74
7.2.3 算法步骤	75
7.2.4 仿真研究	75
7.3 本章小结	78
参考文献	79
第 8 章 基于支持向量机的强化学习	80
8.1 SVM	80
8.1.1 机器学习	80
8.1.2 核学习	82
8.1.3 SVM 的思想	82
8.1.4 SVM 的重要概念	83
8.2 基于 SVM 的强化学习	84
8.2.1 基于 SVM 的 Q 学习结构	84
8.2.2 基于滚动时间窗机制的 SVM	86
8.2.3 算法步骤	87
8.2.4 仿真研究	88
8.3 基于协同最小二乘 SVM 的强化学习	89
8.3.1 基于协同最小二乘 SVM 的 Q 学习	90
8.3.2 LS-SVRM 逼近状态-动作对到值函数的映射关系	92
8.3.3 LS-SVCM 逼近状态空间到动作空间的映射关系	93
8.3.4 仿真研究	93
8.4 本章小结	96
参考文献	96

第 9 章 基于高斯过程分类器的强化学习	98
9.1 基于高斯过程分类器的强化学习	98
9.2 在线高斯过程分类器学习	99
9.3 算法步骤	101
9.4 仿真研究	102
9.5 本章小结	105
参考文献	106
第 10 章 基于图上测地高斯基函数的策略迭代强化学习	107
10.1 环境的图论描述	108
10.2 测地高斯基函数	109
10.3 递归最小二乘策略迭代	111
10.4 算法步骤	112
10.5 仿真研究	112
10.6 本章小结	114
参考文献	115
第 11 章 多目标优化问题的差分进化-分布估计算法	116
11.1 多目标优化	117
11.2 多目标优化的差分进化-分布估计算法	117
11.2.1 多目标优化的 DE-EDA 混合算法步骤	118
11.2.2 多目标优化的 DE 子代生成策略	119
11.2.3 多目标优化的 EDA 子代生成策略	121
11.3 实例研究	121
11.4 本章小结	130
参考文献	130
第 12 章 基于细菌觅食行为的分布估计算法在预测控制中的应用	132
12.1 方法的提出	133
12.2 基于改进分布估计算法的预测控制	133
12.2.1 预测模型	133
12.2.2 反馈校正模型	134
12.2.3 基于改进分布估计算法的滚动优化	134
12.3 实验分析	136
12.3.1 Benchmark 函数实验	136
12.3.2 预测控制的曲线跟踪实验	139
12.4 本章小结	140
参考文献	141

第 13 章 一种多样性保持的分布估计算法	142
13.1 混沌模型	143
13.2 多样性保持分布估计算法	143
13.3 Benchmark 函数实验	146
13.4 在支持向量机参数选择中的应用	152
13.4.1 算法步骤	152
13.4.2 Chebyshev 混沌时间序列预测	153
13.5 本章小结	157
参考文献	157
附录 部分机器学习算法 MATLAB 源程序	159
程序 1 第 11 章 多目标差分进化-分布估计算法 MATLAB 源程序	159
程序 2 第 12 章 基于细菌觅食行为的分布估计算法部分 MATLAB 源 程序	166
程序 3 第 13 章 一种多样性保持的分布估计算法部分 MATLAB 程序	172

第 1 章 机器学习概述

知识、知识表示及运用知识的推理算法是人工智能的核心,而机器学习则是关键问题。随着机器学习研究的不断深入开展和计算机技术的进步,已经设计出不少具有优良性能的机器学习系统并投入实际应用。这些应用领域涉及图像处理、模式识别、机器人动力学与控制、自动控制、自然语言理解、语音识别、信号处理和专家系统等。本章将简要介绍机器学习的概念、发展历史、分类及其部分主要策略。

1.1 机器学习的概念

学习是人类具有的一种重要智能行为^[1],是人们习以为常的一个概念,但究竟什么是学习,长期以来却众说纷纭。社会学家、逻辑学家和心理学家都各有其不同的看法,至今仍无一个统一的定义。产生这一现象的原因主要与以下三个方面的因素有关^[2, 3]:第一,学习是一种综合性的心理活动,它与记忆、思维、知觉、感觉等多种心理活动密切联系,使得人们难以弄清它的机理;第二,学习是一种具有多侧面的现象,它包括获取或发现新知识、修正或精化老知识、归纳或演绎原有知识等多种实践活动,使得人们很难把握它的实质;第三,从事学习研究的学者多来自不同的学科,如社会学家、逻辑学家、心理学家等,他们分别从不同的角度对学习给出了不同的解释、提出了不同的观点。

- ◇ 心理学中对学习的解释:学习是指(人或动物)依靠经验的获得而使行为持久变化的过程。
- ◇ Simon 认为:如果一个系统能够通过执行某种过程而改进它的性能,这就是学习。
- ◇ Minsky 认为:学习是在人们头脑中(心理内部)进行有用的变化。
- ◇ Mitchell 在 *Machine Learning* 一书中对学习的定义:对于某类任务 T 和性能度 P ,如果一个计算机程序在 T 上以 P 衡量的性能随着经验 E 而自我完善,那么,我们称这个计算机程序从经验 E 中学习。
- ◇ 当前关于机器学习的许多文献中也大都认为:学习是系统积累经验以改善其自身性能的过程。

上述观点虽然不尽相同,但却都包含了知识获取和能力改善这两个主要方面。

所谓知识获取是指获得知识、积累经验、发现规律等；所谓能力改善是指改进性能、适应环境、实现自我完善等。在学习过程中，知识获取与能力改善是密切相关的，知识获取是学习的核心，能力改善是学习的结果。

总之：① 学习与经验有关；② 学习可以改善系统性能；③ 学习是一个有反馈的信息处理与控制过程。因为经验是在系统与环境的交互过程中产生的，而经验中应该包含系统输入、响应和效果等信息。因此经验的积累、性能的完善正是通过重复这一过程而实现的。

通过以上分析，我们可以对学习给出如下较为一般的解释：学习是一个有特定目的的知识获取和能力增长过程，其内在行为是获得知识、积累经验、发现规律等，其外部表现是改进性能、适应环境、实现自我完善等。

那么，什么是机器学习呢？它是人工智能中最具有智能特征的前沿研究领域之一，也是知识发现、数据挖掘等领域的重要基础。顾名思义，机器学习(machine learning, ML)是一门研究怎样用计算机来模拟或实现人类学习活动的学科，它是下述多门学科的交叉：

- ◇ 计算机科学：人工智能，理论计算机科学，尤其是计算复杂性理论。
- ◇ 数学：概率和数理统计，信息科学，控制理论。
- ◇ 心理学：人类问题求解和记忆模型。
- ◇ 生物学/遗传学：遗传算法，连接主义。
- ◇ 哲学：Occam 剃刀。

图 1.1 给出了机器学习的简单模型^[4~6]。模型中包含学习系统的四个基本环节。环境和知识库是以某种知识表示形式表达的信息的集合，学习环节和执行环节代表两个过程。环境提供外界信息，类似教师的角色。学习环节处理环境提供的信息，相当于各种学习算法，以便改善知识库中的显式知识。知识库中以某种知识表示形式存储信息。执行环节利用知识库中的知识来完成某种任务，并把执行中的情况回送给学习环节。学习使系统的性能得到改善。机器学习的研究可以使机器自动获取知识，赋予机器更多的智能。另一方面可以进一步揭示人类思维规律和学习奥秘，帮助人们提高学习效率。机器学习的研究还会对记忆存储模式、信息输入方式及计算机体系结构产生重大影响。

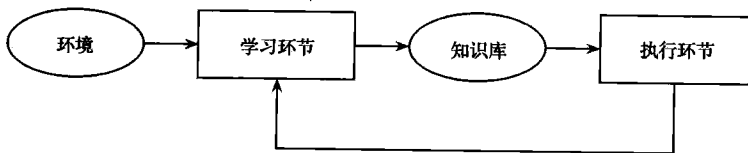


图 1.1 机器学习模型

1.2 机器学习的发展历史

自从20世纪50年代开始研究机器学习以来,在不同时期的研究途径和目标也不同,可以划分为四个阶段^[7]。

第一阶段是20世纪50年代中叶到60年代中叶,属于热烈时期。在这个时期,所研究的是“没有知识”的学习,即“无知”学习,它的主要研究目标是各类自组织系统和自适应系统。例如,如果给系统一组刺激、一个反馈源以及修改它们自身组织的足够自由度,那么它们将改变自身成为最优的组织,即它们能够修改自身以适应它们的环境。这类系统采用的主要研究方法是不断修改系统的控制参数和改进系统的执行能力,不涉及与具体任务有关的知识。本阶段的代表性工作是:Samuel的下棋程序。但这种学习的结果远不能满足人们对机器学习系统的期望。

第二阶段是在20世纪60年代中叶到70年代中叶,被称为机器学习的冷静时期。本阶段的研究目标是模拟人类的概念学习过程,并采用逻辑结构或图结构作为机器内部描述。机器采用符号来表示概念,又称符号概念获取,并提出关于所学概念的各种假设。在这一阶段,研究者意识到学习是复杂而困难的过程,因此人们不能期望学习系统可以从没用任何知识的环境中开始,学习到高深而有价值的概念。这种观点使得研究人员一方面深入探讨简单的学习问题,另一方面则把大量的领域专家知识加入到学习系统中。本阶段的代表性工作有Winston的结构学习系统和Hayes-Roth等的基于逻辑的归纳学习系统。

第三阶段从20世纪70年代中叶到80年代中叶,称为复兴时期。在此期间,人们从学习单个概念扩展到学习多个概念,探索不同的学习策略和学习方法,且在本阶段已开始把学习系统与各种应用结合起来,并取得很大的成功。同时,专家系统在知识获取方面的需求,也极大地刺激了机器学习的研究和发展。在出现第一个专家学习系统之后,示例归纳学习系统成为研究的主流,自动知识获取成为机器学习应用的研究目标。1980年,在美国的卡内基-梅隆(CMU)召开了第一届机器学习国际研讨会,标志着机器学习研究已在全世界兴起。此后,机器学习开始得到了大量的应用。*Strategic Analysis and Information System* 国际杂志连续三期刊登有关机器学习的文章。1984年,由Simon等20多位人工智能专家共同撰文编写的*Machine Learning* 文集第二卷出版,国际性杂志*Machine Learning* 创刊,更加显示出机器学习突飞猛进的发展趋势。这一阶段代表性的工作有Mostow的指导式学习,Lenat的数学概念发现程序,Langley的BACON程序及其改进程序。

第四阶段从20世纪80年代中叶到现在,是机器的最新阶段。这个时期的机器学习具有如下特点:机器学习已成为新的边缘学科,它综合应用了心理学、生物学和神经生理学以及数学、自动化和计算机科学形成机器学习理论基础;融合

了各种学习方法,且形式多样的集成学习系统研究正在兴起;机器学习与人工智能各种基础问题的统一性观点正在形成;各种学习方法的应用范围不断扩大,一部分应用研究成果已转化为商品;与机器学习有关的学术活动空前活跃。

1.3 机器学习的分类

几十年来,研究发表的机器学习的方法种类很多,根据强调侧面的不同可以有多种分类方法^[8]。

1.3.1 基于学习策略的分类

(1) 模拟人脑的机器学习。

◇ 符号学习:模拟人脑的宏观心理级学习过程,以认知心理学原理为基础,以符号数据为输入,以符号运算为方法,用推理过程在图或状态空间中搜索,学习的目标为概念或规则等。符号学习的典型方法有:记忆学习、示例学习、演绎学习、类比学习、解释学习等。

◇ 神经网络学习(或连接学习):模拟人脑的微观生理级学习过程,以脑和神经科学原理为基础,以人工神经网络为函数结构模型,以数值数据为输入,以数值运算为方法,用迭代过程在系数向量空间中搜索,学习的目标为函数。典型的连接学习有权值修正学习、拓扑结构学习。

(2) 直接采用数学方法的机器学习:主要有统计机器学习。

1.3.2 基于学习方法的分类

(1) 归纳学习。

◇ 符号归纳学习:典型的符号归纳学习有示例学习、决策树学习。

◇ 函数归纳学习(发现学习):典型的函数归纳学习有神经网络学习、示例学习、发现学习、统计学习。

(2) 演绎学习。

(3) 类比学习:典型的类比学习有案例(范例)学习。

(4) 分析学习:典型的分析学习有解释学习、宏操作学习。

1.3.3 基于学习方式的分类

(1) 监督学习(有导师学习):输入数据中有导师信号,以概率函数、代数函数或人工神经网络为基函数模型,采用迭代计算方法,学习结果为函数。

(2) 无监督学习(无导师学习):输入数据中无导师信号,采用聚类方法,学习结果为类别。典型的无导师学习有发现学习、聚类、竞争学习等。